THE

UNITED STATES

NATIONAL DATA CENTER

BY

Bruce R. Varnum

September 1995

Directorate of Nuclear Treaty Monitoring

Air Force Technical Applications Center

1030 S. Highway A1A

Patrick Air Force Base, Florida 32925-3002

1.0 INTRODUCTION

The United Nations Conference on Disarmament *ad hoc* Group of Scientific Experts (GSE) is conducting the third large-scale technical test , termed GSETT-3, which started on January 1st of this year. The purpose is the continuation of an on-going effort to develop a global system for international monitoring of a Comprehensive Test Ban Treaty (CTBT) which will provide member states with data for their own national monitoring assessment. Accomplishment of this objective requires testing of a system composed of four major components: an International Data Center (IDC), National Data Centers (NDCs) in each member state, a network of seismic stations in each member state, and high speed communications to share the data in near-real time. This paper will discuss the progress of the US efforts to prototype and operate its NDC for GSETT-3. With eight months having elapsed since GSETT-3 began, we are now in a position to attempt a more consistent statistical assessment of the reliability and capability of the US NDC.

2.0 US NATIONAL DATA CENTER

Under the GSE concept for an international monitoring system, each participating member state has several options from which they may choose to develop their own NDC. Member states may choose to cooperate with one or more other states to form a "Regional" Data Center or a regional data concentrator facility; to develop only a raw data transmission capability; or to develop a full-scale NDC. A full-scale NDC, by definition, contains the capability to transmit and process the data, report results to their own national command authorities, and provide a long-term archive of the data. The United States is participating in GSETT-3 by prototyping and operating a full-scale NDC.

The US NDC for GSETT-3 has been established at the Air Force Technical Applications Center (AFTAC) Patrick Air Force Base, Florida. The United States has elected to prototype its NDC with the capability to: fully manage the US network of GSE seismic stations; submit the data from these stations to the IDC; fully process all data received from the IDC; and incorporate data from other AFTAC sources into its database.

The US NDC is currently receiving, and transmitting to the IDC, data from seven US Primary stations, data from eleven US Auxiliary stations, and data from seven southern hemisphere stations. In addition, Supplemental data are being collected and transmitted to the IDC.

2.1 SEISMIC STATIONS

Eight months into the GESETT-3 experiment finds the US NDC receiving data from the seven US planned Primary stations, the seven Southern hemisphere stations, and eleven of the twelve planned Auxiliary stations. These stations were selected for the following rea-

sons: they had the appropriate hardware configuration, they were situated near areas of high seismicity in the US, and they all had low levels of seismic background noise.

2.1.1 Primary Stations (Alpha)

Currently, there are no direct links from any primary stations to the IDC. All Primary data pass through a NDC or a data relay center. For example, CMAR in Thailand and a number of southern hemisphere stations transmit data to the IDC via a relay co-located with the US NDC. Several arrays located in Europe are transmitted through a similar center located in Norway. Most of the links to the IDC, the United States is an example, carry data from several stations located in that particular country. The US NDC is probably the most complex NDC due to its receiving data from the data sources discussed below. Two of the seven US primary stations are arrays, they are TXAR and PDAR. These stations have been operating for the full eight months with some problems encountered and upgrades incorporated. The remaining Primary stations are 3-component stations, they are: Lisbon, New Hampshire (LBNH); Mount Ida, Arkansas (MIAR); Pinon Flat, California (PFO); North Pole, Alaska (NPO); and Vanda, Antarctica (VNDA). The Southern hemisphere stations are sending 3-component data to the US NDC which is then forwarded to the IDC. These stations are: Brasilia, Brazil (BDFB); Bangui, Central African Republic (BGCA); Boshof, South Africa (BOSA); Villa Florida, Paraguay (CPUP); Dimbroko, Ivory Coast (DBIC); La Paz, Boliva (LPAZ); Paso Flores, Argentina (PLCA) and the only array station Chiang Mai, Thailand (CMAR).

2.1.2 Auxiliary stations (Beta)

The following twelve stations were selected using criteria similar to the Primary stations and they are: Albuquerque, New Mexico (ALQ); Blacksburg, Virginia (BLA); Dugway, Utah (DUG); Elko, Nevada (ELK); Ely, Minnesota (EYMN); Kanab, Utah (KNB); Mina, Nevada (MNV); Newport, Washington (NEW); Black Hills, South Dakota (RSSD); Tucson, Arizona (TUC); Tulsa, Oklahoma (TUL); and Tuckaleechee, Tennessee (TLK). Eleven of these stations (all except KNB) are now sending data routinely to the US NDC. Thus, the US NDC's experience with Auxiliary stations is becoming quite extensive, which is also true for a number of the other countries.

2.2 US NATIONAL DATA CENTER ARCHIVES

2.2.1 Overview

The National Data Center Archives (NDCA) is part of the NDC. The NDCA contains both near real-time continuous raw waveform data and historical collections of segmented data organized around selected events, areas, and/or research projects. For the near future-the largest contributor to the NDCA will be raw seismic waveform data being collected during GSETT-3. The NDCA is scheduled to begin archiving all raw GSETT-3 data by September 1995 with an anticipated flow rate of 2-3 Gbytes/day. Ultimately twelve months worth of data will be on-line (or near on-line). It is anticipated that requests for data will be fullfilled within minutes after the system becomes fully operational. After

the twelve months the older continuous data will be migrated off for permanent storage on DAT tapes while segmented event data will be archived on CD-ROM or optical disks.

2.2.2 Scope of the Data Archive

For planning purposes, the maximum daily flow rate to the NDCA is set at 15 Gbytes/day (5.5 terabytes/year). In addition, historical collections of unclassified seismic, and possibly other types, will be incorporated into the NDCA and transferred to optical disk or CD-ROM. These data will not have an additional quality control function performed on them prior to their inclusion.

2.2.3 Data Definition

All seismic, hydroacoustic, and infrasonic data available through the NDCA are in CSS3.0 format as defined in the Technical Report C90-01 of the Center for Seismic Studies. Station parameters include annotations as to the type of geodetic survey and reference spheroid used to locate the instrument sites. The data supplied to the NDCA includes the instrument response, calibration details, and other station specific information so that this information is available to the user.

Data collected as part of the GSETT-3 experiment for primary and auxillary seismic stations will be transmitted to the US NDC, starting in September, using the IDC-defined alpha protocols. These data will be transmitted to the NDCA as close to real time as possible (within 36 hours of the event times). Primary station data will be recorded and archived continuously, while auxiliary station data may be recorded and archived only for specifically requested segments. The NDCA plans to collect and archive continuous data for all auxiliary stations located in the United States.

2.3 DATA DISTRIBUTION

2.3.1 Types of Requests

The anticipation is that there will be two basic types of requests for raw waveform data. The first is for discrete events recorded at all or selected stations and bounded by a maximum of 100 MBytes per request. The second is a global request for all data received for a period extending for minutes, hours, or continuous current data.

The hardware and software being implemented in the US NDC was designed to handle the expected initial exponential growth. Discrete event requests for on-line data may be retrieved directly by the requestor. All other requests will require attention by the US NDC staff and, in general, discrete event-oriented requests will be given priority over global requests. A fully implemented autoDRM is essential to the successful operation of this data distribution scheme and is scheduled to be operational in October.

2.3.2 Request Format

All data must be requested via the GSE2.0 Message submitted by e-mail to a specific account at the US NDC. Users will be able to browse the archive using the World Wide Web Server (web browsers such as MOSAIC, VIDIA, or Netscape will work), and use this information to prepare GSE2.0 message format data requests. Basic Web browsing and basic autoDRM capability are to be installed in September 1995. Except for extenuating circumstances, users will be expected to retrieve "static" files such as station locations, type of instrumentation, and instrument response curves without assistance from the NDC staff. These static files will be maintained on-line and updated daily using data from the IDC. The static file information can be retrieved by downloading from the Web server or by submitting an autoDRM request.

2.3.3 Request Handling Priority

The US NDC is responsible (first priority) for satisfying requests submitted to it by the IDC for US auxiliary and other data. Official government agency requests will be assigned second priority. Small requests from others agencies and individuals will be assigned third priority Very large requests from this last group of requestors will be assigned the lowest priority.

2.3.4 Expected Request Response Time

With retrieval capability currently available at the US NDC, daily subscriptions to IDC products such as the Reviewed Event Bulletin (REB) is available. Authorized government agencies and contractors who require access to continuous waveforms will need to make special arrangements to access the data. Full implementation of the autoDRM and the World Wide Web Server capabilities is planned for January 1996. The goal is to respond to and fulfill all requests less than 100 MBytes of data within a few days of receipt of request.

2.3.5 Security

Data requestors coming in from the Internet will not have accounts on the system, but will be able to browse the archive contents using the World Wide Web Server, submit e-mail GSE2.0 messages , and be able to retrieve data using anonymous ftp.

2.4 US NDC DATA PROCESSING

AFTAC is continuing the development effort for the US NDC to process the data for treaty monitoring purposes. AFTAC has gone on-line earlier this year with the AFTAC Digital Seismic Network (ADSN). The US NDC processing is relying heavily on the ADSN processing. This processing includes the data handling (i.e. array processing) the signal detection process, the signal association process, the estimation process for magnitude, depth process, etc., and the identification processing which relies on processes such as long period (lp) signal processing. The processing structure has its roots in the ADSN

structure, and ultimately in the IDC structure. With the recent reception of some IDC data, the US NDC will begin preliminary processing of those received data during this month (September 1995).

The final products of the ADSN and the US NDC will be different in the final analysis primarily due the end results that are desired. Basically, the NDC exists to support the treaty verification process while the ADSN has added additional features for in-country specific requirements.

2.4.1 Processing Enhancements

The future enhancements relate to the following: the use of regional discriminants, the expanded use of long period data (location for example), the incorporation of more robust association algorithms, and improved signal detectors. Finally, the addition of additional types of data, discussed in the next section, properly fused with the seismic data and processing should improve the results in certain areas if not overall.

2.5 ADDITIONAL DATA SOURCES

The US NDC is currently acquiring additional data that will be sent to the IDC assuming the agreements are put-in-place for GSETT-3. These additional data are: Hydroacoustic, Radionuclide, and Infrasonic. At present these data are being archived at the US NDC, and will be available later in the experiment.

2.5.1 Hydroacoustic

The US will supply supply data from two hydroacoustic arrays located at Ascension Island and Wake Island. Four additional fixed arrays are planned for installation and the data will also be made available. When the pending agreements are put-in-place all these data will be forwarded to the IDC.

2.5.2 Radionuclide

The US has installed two radionuclide monitoring systems at US facilities. These systems are co-located with the seismic stations at Pinedale, Wyoming and North Pole, Alaska. If the GSE agrees that the IDC should experiment with the non-seismic data these data will also be forwarded to the IDC.

2.5.3 Infrasonic

Under a cooperative program with the US Department of Energy, Los Alamos National Laboratory (LANL) to provide infrasonic data from two arrays, the US NDC is currently acquiring and archiving these data. Again, if the GSE agrees that the IDC should experiment with non-seismic data they will be forwarded to the IDC.

3.0 PERFORMANCE SUMMARY

3.1 Contributions to the IDC by the US NDC

Tables and/or graphs are available for the Primary station contributions to the IDC. These graphs show the daily percentage, of the total amount possible, that data are received at the IDC. These graphs do not break down the specific cause for the data outage. Data are available which gives information on the specific cause for the loss of data. Specificly, the data outage is identified to be caused by software, communication problems, station outages, NDC problems, etc. Unfortunately, these data do not identify the software problems or the communication problems to a particular portion of the system.

The most recent statistics which have been compiled are for the four-month period starting on 1 January and ending on 30 April 1995. A total of 6434 unique events were located and reported in the IDC REB for this time period. The US NDC reported defining phases for 57% of these events. The arrays CMAR, PDAR, and TXAR were among the highest contributors to the IDC REB events, these contributions ranged from 23% to 45% of the IDC REB event total. In addition, NPO and BGCA, both 3-component stations, also contributed 30% and 21% of the IDC REB events respectively.

Preliminary results for this four month period indicate the majority of the US NDC stations significantly contribute to associated events located at teleseismic distances from them (epicenters between 20 and 100 degrees). The Southern Hemisphere stations contributed disproportionately high to the events located in the Northern Hemisphere.

Considering all of the US NDC stations, only PDAR and NPO are contributing to regional events. The other stations with high percentages of regional detections, such as CMAR, TXAR, and CPUP, are contributing detections primarily to teleseismic events.

3.1.1 Experience with the IDC Products

In order to further evaluate the US NDC performance, the US NDC personnel subscribe to the IDC bulletins (AEL, ABEL, and REB) as well as the Alpha-status and Comm-status reports. The principal comparison that has been made to date is the United States located events of the REB with the Supplementary Event List (SELs) which is derived from the USGS weekly Preliminary Determination of Epicenters (PDE) bulletins. This bulletin is made up of events occurring only in the US. Thus, the US located events are compared with the SEL events for the same time period. The comparison generally takes into account the basic differences such as station coverage and the deletion of suspected mining blasts from the SEL The comparison of the SEL and REB events was separated into three categories which are the following: events reported by the SEL but not the REB, events reported by both the SEL and the REB, and events reported by the REB but not the SEL.

4.0 CONCLUSIONS

Two years have passed since AFTAC began development of the US NDC. More recently, the US NDC has been supporting the GSETT-3 experiment for over eight months. The NDC is supplying to the NDC nearly all station data which were originally agreed upon. Data outages continue to be a problem with a wide range of causes. Now that nearly all of the Primary and Auxiliary stations are on-line, more effort can be allocated to locating and resolving these problems. In addition, more hardware and software are being brought on-line, as well as upgrading that which has been functioning, lending credence to the reliability numbers originally proposed in the program.

The next four months will be very important as the IDC data will be increasingly available to the US NDC for archiving and processing. This expansion in the amount of data and number of stations will greatly aid the NDC personnel in optimizing the processing algorithms currently available and those scheduled for near-term inclusion. Also the capability for researchers to request and receive large amounts of data from the US NDC will be made available during this time frame.